INVESTIGATING THE CAPABILITIES OF CFD-BASED DATA-DRIVEN MODELS FOR INDOOR ENVIRONMENTAL DESIGN AND CONTROL

NINA MOROZOVA, F. XAVIER TRIAS, ROSER CAPDEVILA, AND ASSENSI OLIVA

Heat and Mass Transfer Technological Center (CTTC),
Universitat Politècnica de Catalunya - BarcelonaTech (UPC)
ESEIAAT, C/ Colom 11, 08222 Terrassa (Barcelona), Spain
nina.morozova@upc.edu, francesc.xavier.trias@upc.edu,
roser.capdevila@upc.edu, asensio.oliva@upc.edu
http://www.cttc.upc.edu/

Key words: Computational Fluid Dynamics, Data Driven Models, Indoor Environmental Applications, Mixed Convection, Turbulence Modeling

Abstract. In this work, we study the accuracy of CFD-based data-driven models, which predict comfort-related flow parameters in a ventilated cavity with a heated floor. We compare the computational cost and accuracy of three different models, namely artificial neural network, support vector regression, and gradient boosting regression. The tested scenarios include short and long cavities with different inlet velocities. Among the studied frameworks, the artificial neural network provides the most accurate predictions for most of the tested flow configurations. However, test configurations with jet separation and a secondary vortex are more difficult to predict correctly; thus more high-fidelity data is required in order to construct a more robust and reliable model.

1 INTRODUCTION

Heating ventilation and air conditioning (HVAC) systems account for approximately 40% of the energy consumption in buildings [1]. The level of energy consumption can be decreased by proper design and precise control of indoor air parameters. Therefore, the fast and accurate computation of indoor airflow is required for testing different design options or performing model predictive control (MPC). Nowadays, air distribution in buildings is usually evaluated by multizone models [2], zonal models [3], and CFD. Multizone models are the most popular choice due to the low computational cost, but they have limited applicability because each room is represented by only one node. Zonal models are considered intermediate between multizone and CFD, but they usually suffer from case dependency. On the other hand, CFD provides full information about the flow field of the domain. CFD simulations are the most accurate way of simulating indoor airflows, but they require big computational resources.

CFD has been successfully used for detailed simulations of building airflows by many researchers [4, 5]. Nevertheless, all of these studies show the need for a fine grid, thus, big computational resources. Its high computational cost prevents CFD from becoming the primary design tool for building applications, and it is mostly used for research or particular high-end building design. To become a primary design tool, CFD is required to provide sufficient accuracy in capturing flow properties, be fast, and fit into an office computer [6]. Over the last decade, several attempts to reduce the computational cost of CFD for indoor

environmental applications have been made. Wang & Zhai [7] examined the credibility of coarse-grid CFD for HVAC applications and optimized the space discretization to reduce the total truncation error. Kempe & Hantsch [8] applied the immersed boundary method to an LES simulation of a room with a heat source and achieved real-time large-eddy simulations (LES) with 32 CPU cores.

On the other hand, data-driven models (DDMs) are gaining popularity in many industrial applications, including building modeling, due to their accurate approximations of nonlinear processes. For example, Athavale *et al.* [9] and Fang *et al.* [10] compared different DDM approaches for temperature prediction in data centers, and concluded that results produced by the model are in good agreement with the reference CFD results. Warey *et al.*[11] created a vehicle cabin thermal comfort model using machine learning (ML) and high-fidelity CFD simulation results with a test error of less than 5%. Moreover, DDMs are widely used in MPC. For instance, DDM-MPC for HVAC systems were developed for a university building [12], an airport [13], and a residential building [14]. All the aforementioned researchers note that DDMs produce accurate predictions at a low computational cost.

The objective of this work is to develop and analyze ML-based DDM, which uses the data from high-fidelity CFD simulations. The main focus of our research is on investigating the capabilities and limitations of this model as a cheaper alternative to CFD, taking into account specific requirements for indoor environmental applications.

The content of this paper is organized as follows: Section 2 describes the previously conducted analysis of the feasibility of high-fidelity CFD simulations for routine use in building applications, Section 3 describes the details of the test cases and the governing equations, Section 4 describes the numerical methods, Section 5 shows obtained results and their analysis, Section 6 provides a discussion and contains concluding remarks.

2 ANTECEDENTS

In our previous work [6], we studied the feasibility of affordable, fast, and high-fidelity CFD simulations for indoor environmental applications, considering two representative test cases and a wide range of numerical setups. We considered three possible building applications: conceptual design, detailed design, and MPC. To become a primary design tool, CFD is required to provide sufficient accuracy in capturing flow properties, be fast, and fit into an office computer. In most cases, HVAC applications require many fast simulations using limited computational resources in order to be able to test different design options or to predict the optimal MPC strategy. Historically the high computational cost prevents CFD from becoming the primary tool for indoor environmental simulations. Nonetheless, the growth in computational power and advances in numerical algorithms provide an opportunity to use accurate and yet affordable CFD.

We tested LES, unsteady Reynolds averaged Navier-Stokes (URANS), and no-model approaches with both staggered [15] and collocated discretizations [16] on a set of structured Cartesian non-uniform grids. LES and no-model approach showed considerably higher accuracy than URANS. Even though URANS simulations were the least computationally expensive, their accuracy was often insufficient. This could be explained by the nature of the URANS models, as they are not suitable for natural convection and mass separation flows. RANS models were developed mostly for aerodynamic applications, where boundary layers are usually attached or partially attached. Staggered symmetry-preserving discretization considerably improves the accuracy of the results for coarse and extremely coarse grids. In general,

the LES-S3PQ [17] model on staggered grids shows the best trade-off between computational cost and accuracy. However, the required accuracy was achieved only using fine computational grids, which resulted in unreasonably long run-times. We concluded that obtained run-times are too long to make CFD a primary tool for HVAC applications.

Results indicate that fast and reliable CFD simulations on office computers are currently not available for HVAC applications. Moreover, wast computational resources are usually available only for particular high-end building designs and not for routine use. Thus, we estimated the growth of computational resources in the future to determine when CFD would be available for routine use on office computers. We adopted Moore's law [18] as an optimistic prediction and the bandwidth growth rate as a pessimistic one. The optimistic prediction estimates that CFD would be feasible for conceptual design in 5 years, for the detailed design in 10-20 years, and 10-25 years for MPC. The pessimistic prediction anticipates at least 15 years for the conceptual design, 25 years for the detailed design, and 35 years for control. The growth of computational resources would not be enough to make CFD available for routine use in building applications in the foreseeable future. This means more work is required on developing better models and numerical methods to reduce the computational cost of the simulations while maintaining accuracy.

3 GOVERNING EQUATIONS AND PHYSICAL PROBLEM

The incompressible Navier-Stokes equations for a Newtonian fluid with constant physical properties are considered. The Boussinesq approximation is adopted to account for the density variations due to temperature difference. Thermal radiation is neglected. Under these assumptions, the governing equations are

$$\nabla \cdot \mathbf{u} = 0, \tag{1}$$

$$\frac{\partial \mathbf{u}}{\partial t} + (\mathbf{u} \cdot \nabla)\mathbf{u} = \nu\nabla^2\mathbf{u} - \nabla p + \beta \mathbf{g} \Delta T, \tag{2}$$

$$\frac{\partial T}{\partial t} + (\mathbf{u} \cdot \nabla)T = \alpha\nabla^2 T, \tag{3}$$

where $\mathbf{u} = (u, v, w)$ is the velocity vector in Cartesian coordinates $\mathbf{x} = (x, y, z)$, p the kinematic pressure, T the temperature, ν the kinematic viscosity, ρ the density, $\mathbf{g}$ the gravitational acceleration, β the thermal expansion coefficient and α the thermal diffusivity.

The physical setup used is a three-dimensional ventilated cavity with a heated floor. This configuration was first studied experimentally by [19]. The geometry of the studied cavity is shown in Figure 1 (left). The cavity has a depth aspect ratio of $A_d = D/H = 0.3$. Cold air at $T_c = -0.5$ enters the cavity through the long thin inlet at the top of the left wall. The inlet velocity profile in the vertical (y) direction corresponds to a parabolic Poiseuille flow with a bulk velocity $U_{in} = 1$. The inlet slot has an aspect ratio $A_{in} = h_{in}/H = 0.018/1.04$. The air is discharged through the outlet with an aspect ratio $A_{out} = h_{out}/H = 0.024/1.04$ at the bottom of the right wall of the cavity. The bottom wall is maintained at a hot temperature of T_h, while the three other sidewalls are kept at the cold temperature of T_c. The cavity is filled with air ($Pr = \nu/\alpha = 0.71$) at Rayleigh number based on the cavity height $Ra_H = \rho g \beta \Delta T H^3/(\nu\alpha) = 2.4 \times 10^9$.

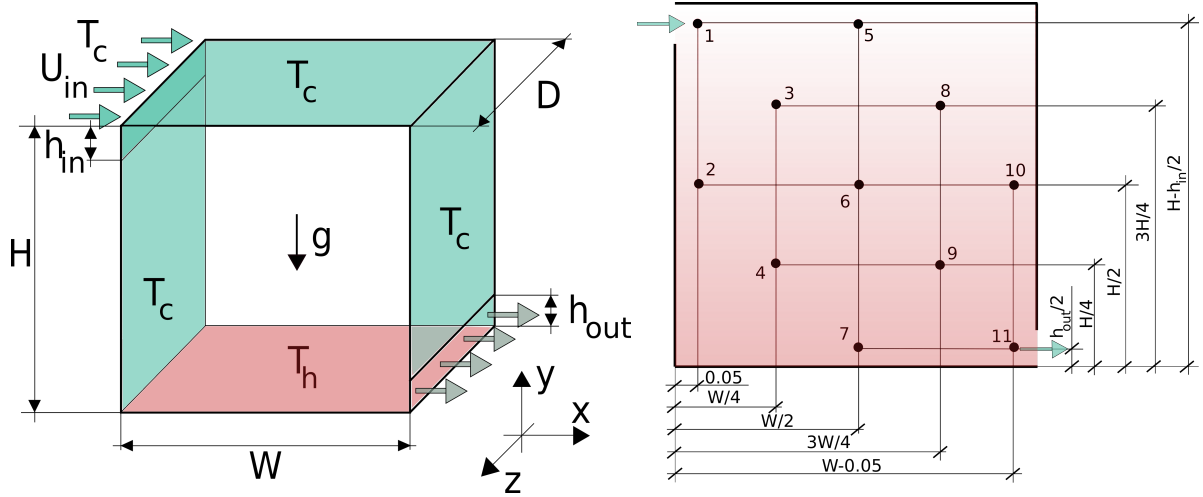All simulations run for 500 non-dimensional time units, which was found to be a long enough time-

Figure 1: Left: geometry of the studied test case. Right: locations of the input data probes.

integration period to record the flow statistics for further averaging. Hereafter, all the results are presented in dimensionless form. The reference values of time, velocity, temperature, and length are $t_{ref} = H/U_{ref}$, $U_{ref} = U_{in}$, $\Delta T = T_h - T_c$, and H, respectively.

We build the input-output dataset by changing the height aspect ratio of the cavity and the Froude number based on the inlet height. We use 5 different height aspect ratios - $A_h = W/H = [0.25, 0.5, 1.0, 2.0, 4.0]$ and 20 different Froude numbers - $Fr_{h_{in}} = U_{in}/\sqrt{\rho g \beta \Delta T h_{in}} = [1.10, 1.50, 2.00, ..., 5.00, 5.24, 5.50, ..., 10.00]$. The changes in the Froude number are produced by variations in the inlet bulk velocity, thus Pr and Ra_H numbers remain constant. These combinations of $Fr_{h_{in}} - A_h$ are realistic and relevant for the indoor environmental applications.

4 NUMERICAL METHODS

4.1 Data generation

Developed DDMs are based on data obtained by high-fidelity CFD simulations using the LES-S3PQ [17] turbulence model with staggered grid [15] discretization. For the cavity with the $A_h = 1$ the computational grid resolution is $N_x \times N_y \times N_z = 100 \times 160 \times 32 = 5.12 \times 10^5$ control volumes. The number of grid points in the horizontal (x) direction is rescaled according to the value of A_h for the other aspect ratios. In total, we carry out 100 CFD simulations, 15% of which are reserved for testing, and 85% are used in the model training.

As input parameters of the model, we consider Froude number ($Fr_{h_{in}}$), cavity height aspect ratio (A_h), temperature (T), and velocity magnitude (V) probes at 11 different locations on the mid-depth cavity plane ($z = D/2$). In total, we use 24 ($Fr_{h_{in}} + A_h + 11T + 11V$) input parameters. The positions of the probes are shown in Figure 1 (right). As the output parameters of the model, we use average Nusselt number on the hot wall - $<Nu>$, jet separation point - x_{sep}, average kinetic energy - $<E>$, average enstrophy - $<\Omega>$, and the average temperature of the cavity. The jet separation point is a point at the top cavity wall, where the wall-shear stress $<\tau_W>$ is equal to zero, as shown in the equation (6). Other flow parameters are calculated using equations (4) - (5).

$$<E> = \frac{1}{V} \int_V \frac{\mathbf{u}^2}{2} dV \qquad <\Omega> = \frac{1}{V} \int_V \omega^2 dV \qquad <T_V> = \frac{1}{V} \int_V T dV \qquad (4)$$

$$<Nu> = -\frac{1}{A} \int_A \frac{\partial T}{\partial x} dA \quad \text{at} \quad x = 0 \qquad (5)$$

$$x_{sep} = x, \quad \text{at} \quad <\tau_W> = 0, \qquad (6)$$

where V is the volume of the cavity, $\omega = \nabla \times \mathbf{u}$ is the vorticity, A is the surface of the hot wall, and $\tau = \frac{\partial u}{\partial y}$ at $y = H$. Developed models use the time-averaged values of the input-output flow parameters.

4.2 DDM frameworks

We use three different ML-based DDM frameworks, namely artificial neural network (ANN), support vector regression (SVR), and gradient boosting regression (GBR).

- ANN is the framework where the output is mapped to the input using a set of interconnected nodes or neurons. In our work, we use the open-source Keras [20] ANN ML library. The neural network used in this work consists of one input layer, one hidden layer, and one output layer. The number of neurons in the hidden layer is equal to 16. Each neuron in the hidden layer is connected to every neuron in the preceding layer via links with specific weights W_{ij}. The output of a neuron y_j in a layer is equal to $y_j = f(\sum_{i=1}^{n} W_{ij} x_i + b j)$, where n_i is the number of neurons in the preceding layer, i is the index for neurons in the preceding layer, f is the non-linear activation function, x_i is the input parameter, and b_j is the bias term associated with the neuron. To improve the results of the prediction and avoid model overfitting, we adopt 10-fold cross-validation [21] and dropout (dropout rate equal to 0.2) regularisation [22] techniques.

- SVR [23] is an ML framework, which encodes non-linear relationships between the inputs and outputs by mapping the data in a high dimensional feature space, where the number of dimensions corresponds to the number of inputs, and performing linear regression in that space. The mapping function is also called the "kernel function". The third-degree polynomial basis kernel function ($K(x, y) = (x^T y + \sigma)^3$) was used in this study, where x and y are vectors in the input space, and σ is a free parameter.

- GBR [24] is a technique, which produces a prediction model that combines weak learners (decision trees) into a single strong learner in an iterative fashion. The algorithm adds one tree at each stage, which optimizes a cost function by moving in the negative gradient direction. For both SVR and GBR frameworks, we used the open-source Scikit-learn [25] library. To improve the results of the prediction, we adopt 10-fold cross-validation [21] for both SVR and GBR models.

5 RESULTS

In this section, we present the prediction results of the different models and analyze how model architecture and the input data affect DDMs accuracy. As a measure of model accuracy, we use relative error (RE), equation (7), and mean relative error (MRE), equation (8). We assume that the less than 15% RE is acceptable for the developed models. The errors are calculated as follows:

$$RE(\phi) = \frac{|\phi_{CFD} - \phi_{DDM}|}{|\phi_{ref}|}, \tag{7}$$

$$MRE(\phi) = \frac{1}{N}\sum_{i=1}^{N} RE(\phi), \tag{8}$$

where N is the number of data points, ϕ_{DDM} stands for the DDM prediction, and ϕ_{CFD} stands for the reference CFD data.

Table 1 shows the MREs of the time-averaged flow parameters predicted by different ML-based DDMs. The value of MRE an average of all of the $Fr_{h_{in}} - A_h$ combinations from the test dataset. The Nusselt number on the hot wall appears to be the most difficult parameter to predict correctly. Even though ANN and SVR models have an acceptable level of accuracy, GBR shows the error as high as 36%. Jet separation point and average cavity temperature were accurately predicted by all of the models. ANN shows the most accurate prediction of kinetic energy and enstrophy. Two other approaches have bigger MREs, which, however, remain close to the acceptable 15%.

Table 1: Mean relative prediction error (MRE) of the studied flow parameters for different models.

Model name	MRE				
	$<Nu>$	x_{sep}	$<T_V>$	$<E>$	$<\Omega>$
ANN	0.144	0.014	0.018	0.086	0.095
GBR	0.357	0.012	0.011	0.166	0.118
SVR	0.132	0.033	0.013	0.149	0.126

The local prediction accuracy is another important aspect of the model performance evaluation. The local accuracy is analyzed using two-dimensional plots, where the local relative prediction error is plotted for the tested combinations of Froude number ($Fr_{h_{in}}$) and height aspect ratios (A_h). Tested combinations are chosen randomly from the dataset and were not used in the model training process. The plots are shown in the Figures 2-6. The horizontal (x) axis of the figures is the Froude number, the spanwise (y) axis is the cavity height aspect ratio, and the vertical (z) axis is the relative error $RE(\phi)$, detailed in the equation (7). The colour scales from light to dark show the values of errors from high to low, respectively.

Relative errors for the Nusselt number are shown in Figure 2. ANN (Figure 2 top left) predicts Nusselt number well for high Froude numbers, but the error is higher for the smaller Froude numbers. Cavities with smaller $Fr_{h_{in}}$ often have a secondary vortex, which makes the correct prediction of $<Nu>$ more difficult. GBR (Figure 2 top right) predicts Nusselt number well for most of the combinations of input parameters, except for two particular peaks in the area of the small cavity height aspect ratio (tall cavity). The high errors could be explained by the dominance of the natural convection in tall cavities, which produces stronger flow stratification. SVR (Figure 2 bottom) exhibits similar to ANN behaviour; however, the overall prediction accuracy is lower.

Flow separation point is shown in the Figure 3. It is predicted accurately for most of the input data combinations by most of the approaches. Even though the tendency is similar for all of the approaches -
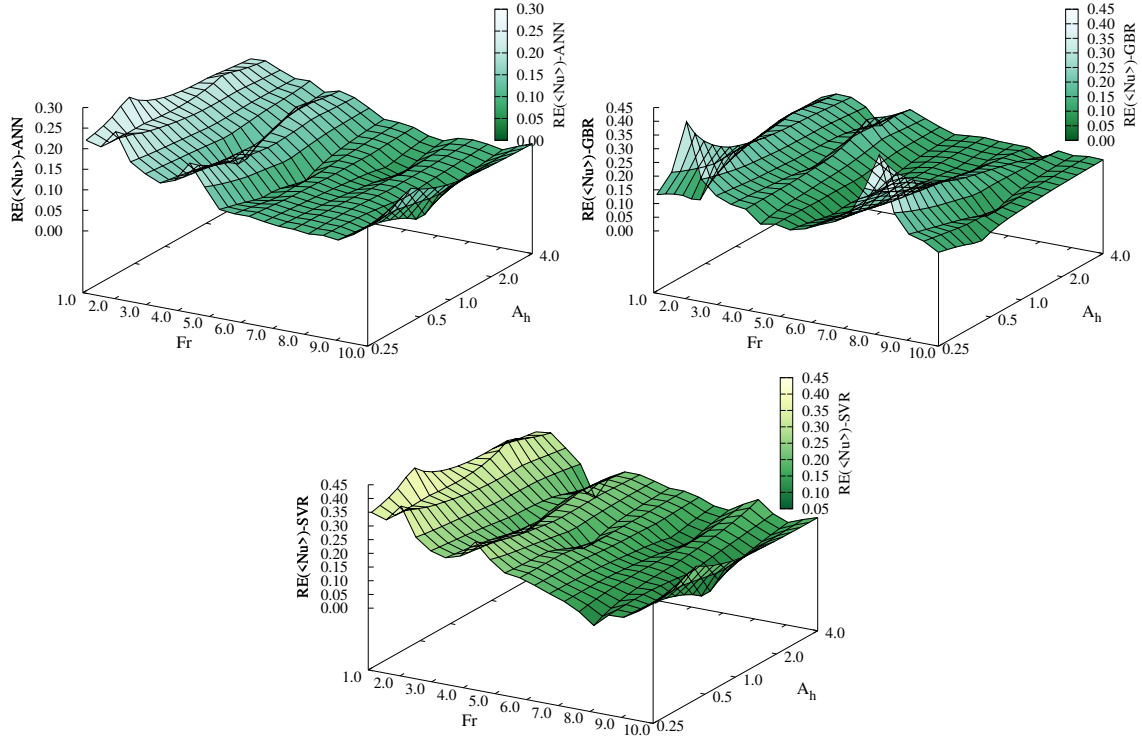
Figure 2: Relative error (RE) of the Nusselt number $<Nu>$ prediction plotted for different combinations of input parameters ($Fr_{h_{in}} - A_h$) and different modeling frameworks.

lower accuracy in the area with small Froude numbers, SVR shows unacceptably high errors in that area. The test configurations with the bigger Froude numbers have the flow separation point closer to the right wall of the cavity or at the right wall itself, which makes its accurate evaluation easier.

The average cavity temperature (Figure 4) is accurately predicted for all of the Froude number - cavity width combinations by all of the approaches. The relative error does not exceed 10%. The tendency is similar to the Nusselt number. The errors are higher in the area with the small Froude numbers, where the natural convection is dominating over the forced one.

Average kinetic energy has an overall high level of relative errors. ANN (Figure 5 top left) gives the most accurate prediction, while GBR (Figure 5 top right) and SVR (Figure 5 bottom) experience high errors peaks in the areas with the smaller Froude numbers. The accuracy of the prediction depends on the level of flow separation. In the areas with a higher $Fr_{h_{in}}$ number, flow separation starts closer to the right wall of the cavity; thus, the secondary vortex is not formed. The presence of the secondary vortex makes the flow more complex hence, the relative error increases.

Average enstrophy results are very similar to the ones of kinetic energy. The areas with the bigger Froude numbers have lower relative errors. ANN (Figure 6 top left) gives the most accurate prediction. GBR (Figure 6 top right) and SVR (Figure 6 bottom) have local areas of high errors, despite the mainly accurate results.

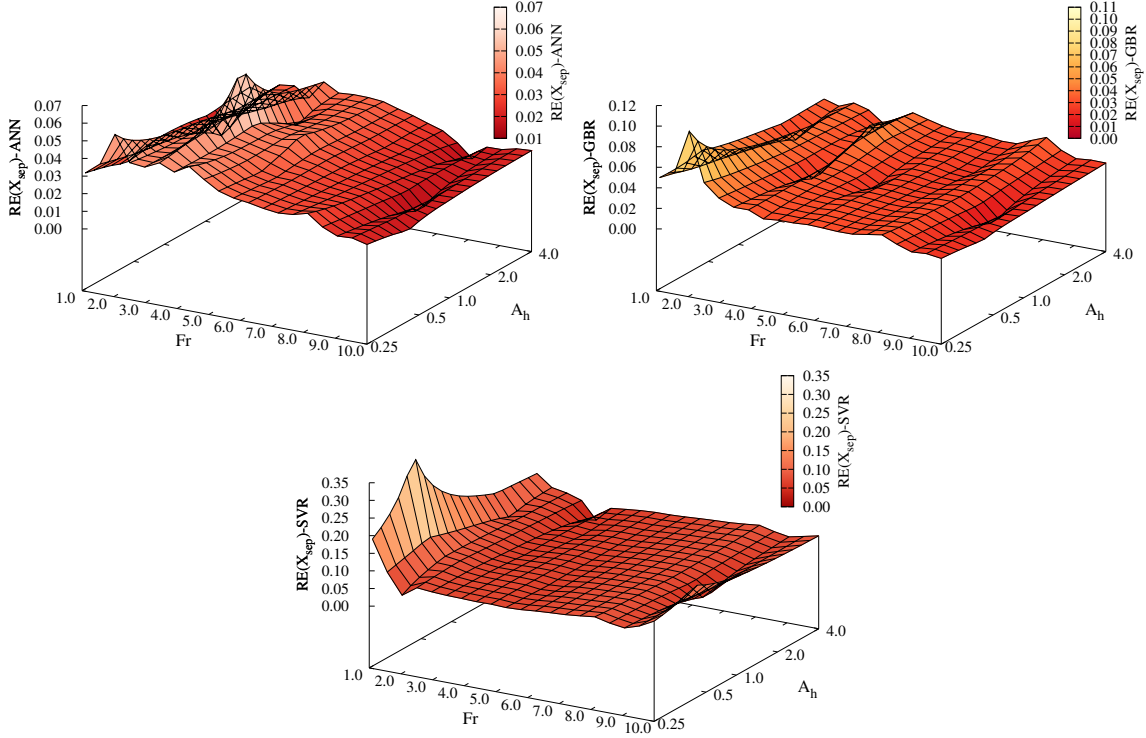The accuracy of the ANN model is rather evenly distributed over the tested $Fr_{h_{in}} - A_h$ configurations.

Figure 3: Relative error (RE) of the flow separation point x_{sep} prediction plotted for different combinations of input parameters ($Fr_{h_{in}} - A_h$) and different modeling frameworks.

Unlike GBR and SVR, which have peaks of local high RE in the area with the smaller Froude numbers. Average cavity temperature and flow separation point are easy to predict accurately. On the other hand, Nusselt number on the hot wall, average kinetic energy, and enstrophy require more training data to improve local accuracy. More data provided for these particularly complicated combinations of inlet velocity and cavity length could potentially improve the quality of the predictions.

6 DISCUSSION AND CONCLUSIONS

In this work, we developed ML-based DDM for predicting the airflow parameters in the ventilated room using three different ML frameworks. The input-output dataset is generated using high-fidelity CFD simulations. The model input parameters are the values of the temperature and velocity magnitude at 11 locations within the cavity domain. The output parameters are the flow parameters relevant to thermal comfort. The developed model is applicable for the flow configurations with complex physical phenomena and commonly used geometry. These flows cannot be accurately resolved by other reduced-order models due to their complexity. On the other hand, the widely used building geometry makes the generation of the input data easier. Hence, DDMs could be used for applications where a combination of fast and accurate predictions is required, for example, for model predictive control.

Among studied modeling frameworks, the artificial neural network showed the most accurate predictions. On the other hand, support vector regression and gradient boosting regression could not correctly predict flow parameters for the test configurations with a jet separation and a secondary vortex. Even though
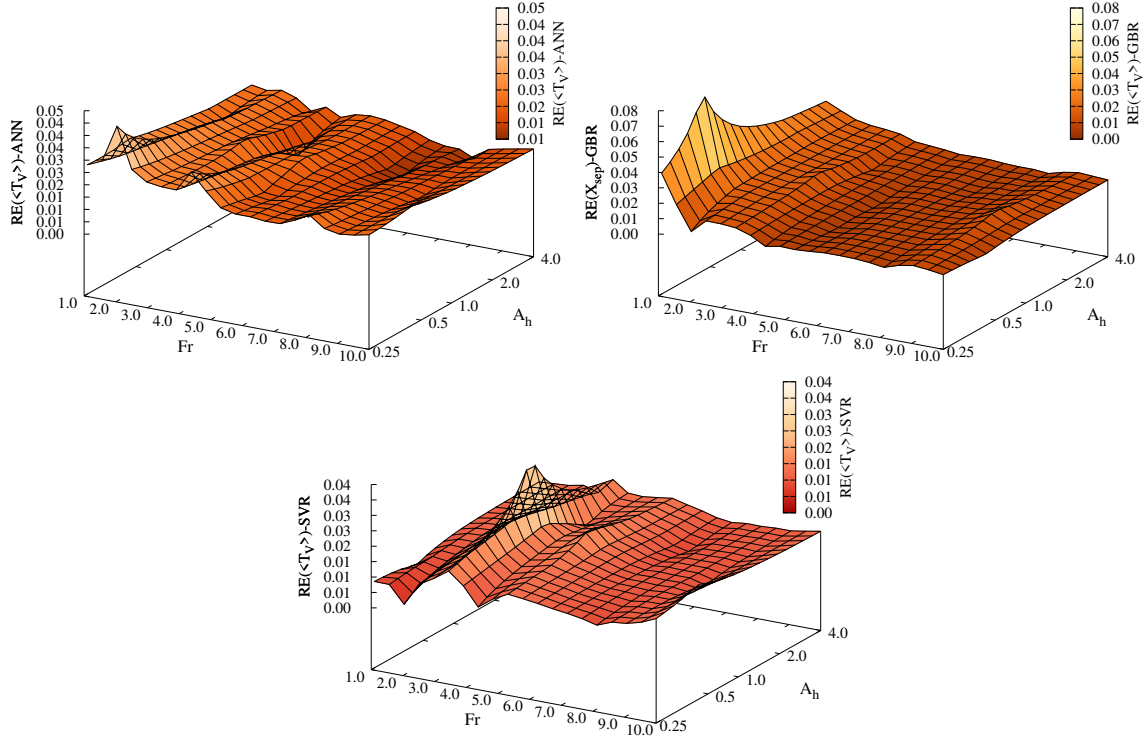
Figure 4: Relative error (RE) of the average temperature $<T_V>$ prediction plotted for different combinations of input parameters $(Fr_{h_{in}} - A_h)$ and different modeling frameworks.

ANN can predict flow parameters with acceptable accuracy for most of the tested flow configurations, more high-fidelity data is required to construct a more robust and reliable model.

The computational cost of a DDM is a combination of the data generation cost and the model training cost. Data generation cost is much higher than the model training cost. The training process is completed within few minutes on an ordinary office computer using only one CPU core. Moreover, after the training is completed, the prediction is produced almost instantly. Even though the computation cost of CFD simulations for data generation is high, these CFD simulations are carried out only once to form the input dataset, and will not be repeated. It takes approximately 215 CPU hours per simulation; thus, the whole dataset with 100 simulations uses around 21500 CPU hours. Taking into account that the estimated electricity cost of one CPU hour using Amazon Web Services [26] is 0.01€, the total electricity cost of the used dataset with 100 CFD simulations is 215€. This is a reasonable computational and financial cost for DDM development.

The work on improving this model is ongoing. In the future, we would like to amplify the available training data by carrying out more-high fidelity LES and RANS simulation, study how different input configurations affect the quality of the predictions using results of direct numerical (DNS) simulations as a reference and find a trade-off between the quantity and the quality of the training data, namely, explore different turbulence models and grid resolutions. Moreover, we plan to work on the extrapolating capabilities of the DDMs.
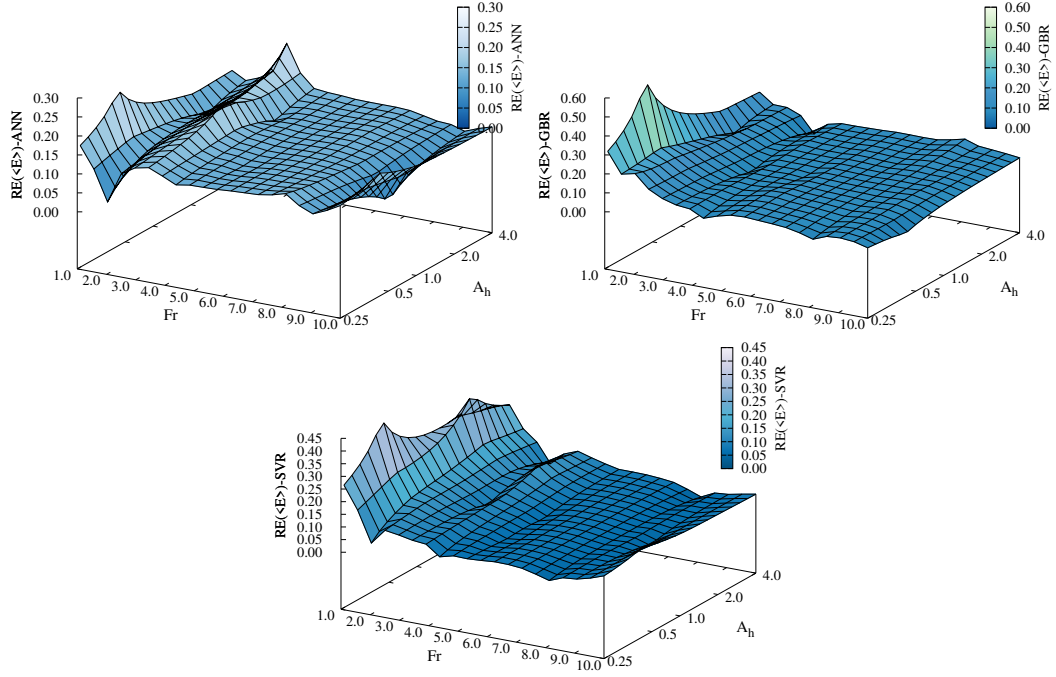
Figure 5: Relative error (RE) of the average kinetic energy $<E>$ prediction plotted for different combinations of input parameters ($Fr_{h_{in}} - A_h$) and different modeling frameworks.
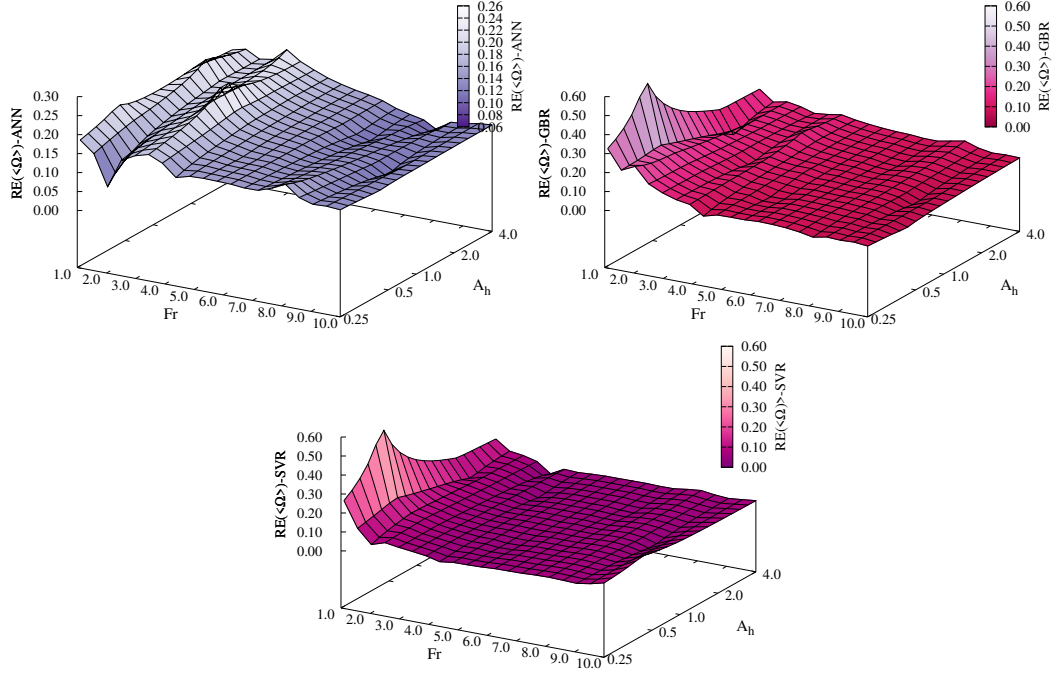

Figure 6: Relative error (RE) of the average enstrophy $<\Omega>$ prediction plotted for different combinations of input parameters ($Fr_{h_{in}} - A_h$) and different modeling frameworks.

ACKNOWLEDGEMENTS

References

[1] X. Cao, X. Dai, and J. Liu, "Building energy-consumption status worldwide and the state-of-the-art technologies for zero-energy buildings during the past decade," *Energy Build.*, vol. 128, pp. 198–213, 2016.

[2] J. Axley, "Multizone airflow modeling in buildings: History and theory," *HVAC&R*, vol. 13, no. 6, pp. 907–928, 2007.

[3] A. C. Megri and H. Fariborz, "Zonal modeling for simulating indoor environment of buildings: Review, recent developments, and applications," *HVAC&R*, vol. 13, no. 6, pp. 887–905, 2007.

[4] T. van Hooff, B. Blocken, and Y. Tominaga, "On the accuracy of CFD simulations of cross-ventilation flows for a generic isolated building: Comparison of RANS, LES and experiments," *Build. Environ.*, vol. 114, pp. 148–165, 2017.

[5] X. Zheng, H. Montazeri, and B. Blocken, "CFD simulations of wind flow and mean surface pressure for buildings with balconies: Comparison of RANS and LES," *Build. Environ.*, vol. 173, p. 106747, 2020.

[6] N. Morozova, F. X. Trias, R. Capdevila, C. D. Pérez-Segarra, and A. Oliva, "On the feasibility of affordable high-fidelity CFD simulations for indoor environment design and control," *Build. Environ.*, vol. 184, p. 107144, 2020.

[7] H. Wang and Z. Zhai, "Application of coarse-grid computational fluid dynamics on indoor environment modeling: Optimizing the trade-off between grid resolution and simulation accuracy," *HVAC&R*, vol. 18, no. 5, pp. 915–933, 2012.

[8] T. Kempe and A. Hantsch, "Large-eddy simulation of indoor air flow using an efficient finite-volume method," *Build. Environ.*, vol. 115, pp. 291–305, 2017.

[9] J. Athavale, M. Yoda, and Y. Joshi, "Comparison of data driven modeling approaches for temperature prediction in data centers," *Int. J. Heat Mass Transf.*, vol. 135, pp. 1039–1052, 2019.

[10] Q. Fang, Z. Li, Y. Wang, M. Song, and J. Wang, "A neural-network enhanced modeling method for real-time evaluation of the temperature distribution in a data center," *Neural Comput. Appl.*, vol. 31, no. 12, pp. 8379–8391, 2019.

[11] A. Warey, S. Kaushik, B. Khalighi, M. Cruse, and G. Venkatesan, "Data-driven prediction of vehicle cabin thermal comfort: using machine learning and high-fidelity simulation results," *Int. J. Heat Mass Transf.*, vol. 148, p. 119083, 2020.

[12] A. E. Ruano, S. Pesteh, S. Silva, H. Duarte, G. Mestre, P. M. Ferreira, H. R. Khosravani, and R. Horta, "The imbpc hvac system: A complete mbpc solution for existing hvac systems," *Energy Build.*, vol. 120, pp. 145 – 158, 2016.

[13] H. Huang, L. Chen, and E. Hu, "A neural network-based multi-zone modelling approach for predictive control system design in commercial buildings," *Energy Build.*, vol. 97, pp. 86–97, 2015.

[14] A. Afram, F. Janabi-Sharifi, A. S. Fung, and K. Raahemifar, "Artificial neural network (ANN) based

model predictive control (MPC) and optimization of HVAC systems: A state of the art review and case study of a residential HVAC system," *Energy Build.*, vol. 141, pp. 96–113, 2017.

[15] R. W. C. P. Verstappen and A. E. P. Veldman, "Symmetry-preserving discretization of turbulent flow," *J. Comput. Phys.*, vol. 187, no. 1, pp. 343–368, 2003.

[16] F. X. Trias, O. Lehmkuhl, A. Oliva, C. D. Pérez-Segarra, and R. W. C. P. Verstappen, "Symmetry-preserving discretization of Navier-Stokes equations on collocated unstructured grids," *J. Comput. Phys.*, vol. 258, pp. 246–267, 2014.

[17] F. X. Trias, D. Folch, A. Gorobets, and A. Oliva, "Building proper invariants for eddy-viscosity subgrid-scale models," *Phys. Fluids*, vol. 27, no. 6, 2015.

[18] G. E. Moore, "Cramming more components onto integrated circuits," *Proceedings of the IEEE*, vol. 86, no. 1, pp. 82–85, 1998.

[19] D. Blay, S. Mergui, J. L. Tuhault, and F. Penot, "Experimental turbulent mixed convection created by confined buoyant wall jets," in *Proceedings of the First European Heat Transfer Conference*, pp. 821–828, 1992. Birmingham (UK), Sept 1992.

[20] F. Chollet, "keras." `https://github.com/fchollet/keras`, 2015.

[21] M. Stone, "Cross-validatory choice and assessment of statistical predictions," *J. R. Stat. Soc. Series B Stat. Methodol.*, vol. 36, no. 2, pp. 111–133, 1974.

[22] N. Srivastava, G. Hinton, A. Krizhevsky, I. Sutskever, and R. Salakhutdinov, "Dropout: A simple way to prevent neural networks from overfitting," *J. Mach. Learn. Res.*, vol. 15, no. 56, pp. 1929–1958, 2014.

[23] C. Cortes and V. Vapnik, "Support-vector networks," in *Machine Learning*, pp. 273–297, 1995.

[24] T. Hastie, R. Tibshirani, and J. Friedman, *The Elements of Statistical Learning*. Springer, 2009.

[25] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay, "Scikit-learn: Machine learning in Python," *J. Mach. Learn. Res.*, vol. 12, pp. 2825–2830, 2011.

[26] Amazon Web Services, "Amazon EC2 On-Demand Pricing." `https://aws.amazon.com/ec2/pricing/on-demand/`, 2021.